

Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*

D. S. Zarlenga^{†‡§}, B. M. Rosenthal[†], G. La Rosa[¶], E. Pozio[¶], and E. P. Hoberg[†]

[†]Bovine Functional Genomics Laboratory and [‡]Animal Parasitic Diseases Laboratory and U.S. National Parasite Collection, U.S. Department of Agriculture, Beltsville, MD 20705; and [¶]Department of Infectious, Parasitic, and Immunomediated Diseases, Istituto Superiore di Sanità, 00161 Rome, Italy

Communicated by William C. Campbell, Drew University, Madison, NJ, March 27, 2006 (received for review December 16, 2005)

Parasitic nematodes of the genus *Trichinella* cause significant food-borne illness and occupy a unique evolutionary position at the base of the phylum Nematoda, unlike the free-living nematode *Caenorhabditis elegans*. Although the forthcoming genome sequence of *Trichinella spiralis* can provide invaluable comparative information about nematode biology, a basic framework for understanding the history of the genus *Trichinella* is needed to maximize its utility. We therefore developed the first robust and comprehensive analysis of the phylogeny and biogeographic history of *Trichinella* using the variation in three genes (nuclear small-subunit rDNA, and second internal transcribed spacer, mitochondrial large-subunit rDNA, and cytochrome oxidase I DNA) from all 11 recognized taxa. We conclude that (i) although Trichinellidae may have diverged from their closest extant relatives during the Paleozoic, all contemporary species of *Trichinella* diversified within the last 20 million years through geographic colonization and pervasive host switching among foraging guilds of obligate carnivores; (ii) mammalian carnivores disseminated encapsulated forms from Eurasia to Africa during the late Miocene and Pliocene, and to the Nearctic across the Bering Land Bridge during the Pliocene and Pleistocene, when crown species ultimately diversified; (iii) the greatest risk to human health is posed by those species retaining an ancestral capacity to parasitize a wide range of hosts; and (iv) early hominids may have first acquired *Trichinella* on the African savannah several million years before swine domestication as their diets shifted from herbivory to facultative carnivory.

biogeography | mitochondrial DNA | phylogeny | ribosomal DNA

Exceptional biological diversity among nematodes is exemplified by certain attributes of the parasite *Trichinella spiralis*. Organization of its mitochondrial genome more closely resembles that of coelomate metazoans than that of its presumed closest relatives, the secernentean nematodes (1). In addition, *T. spiralis* (Dorylaimia) shares a similar proportion (45%) of its ESTs with the nematode *Caenorhabditis elegans* (Rhabditina) as it does with the fruit fly *Drosophila melanogaster* (Arthropoda: Drosophilidae) (2). Thus, many ESTs common to *T. spiralis* and *C. elegans* are not necessarily specific to nematodes but may be conserved among diverse taxonomic groups of invertebrates. *C. elegans* is often thought of as a prototypical nematode because of its acceptance as a model for studying biological processes; however, genomic variation among nematodes is extensive and commensurate with their phylogenetic and ecological diversity. Therefore, the forthcoming genome sequence of *T. spiralis* will contribute substantially to our understanding of nematode biology and the origins of parasitism.

In 1998, Blaxter *et al.* (3) used genetic data to delineate the phylum Nematoda into supertaxa consisting of five clades. *Trichinella*, a parasite of vertebrates, occupies a basal lineage (clade I) consisting of free-living Mononchida, plant parasitic Dorylaimida, and entomophagous Mermithida, with which it shares features of early embryogenesis (4) and small-subunit (SSU) rDNA sequences (3). Other clades show similar levels of

diversity in host-parasite associations, suggesting that parasites of vertebrates arose independent of one another during the long evolutionary history for nematodes (3). Thus, understanding the origins and persistence of vertebrate parasitism requires phylogenetic and historic information from a broad spectrum of taxa. Among the Trichocephalida, parasitism may have arisen as early as the Paleozoic in the ancestor of the Trichinellidae and the Trichuridae; however, temporal origins of extant species of *Trichinella* remain unexplored.

The genus *Trichinella* comprises a monophyletic lineage in the Trichinellidae, the putative sister to the Trichuridae (Capillariinae, Trichurinae, and Trichosomoidinae). The superfamily Trichinelloidea to which *Trichinella* belongs is phylogenetically diagnosed by the stichosome, a region of the glandular esophagus, and the bacillary bands, an assembly of structural characters otherwise unknown among the nematodes. Trichinellidae is further diagnosed by a unique life cycle where worms undergo complete development within a single vertebrate host (autohexeroxy) and by first-stage larvae, which localize within skeletal muscle cells as infectious organisms. Although eating raw or undercooked pork infected with *T. spiralis* accounts for most human cases of trichinellosis, zoonotic transmission also occurs from species that circulate among wildlife. Eleven distinct taxa are now recognized in mammals, birds, crocodilians, and saurians (5), characterized by either a thick (encapsulated) or a very thin (nonencapsulated) collagen membrane enveloping the parasite-infected muscle cell.

Despite extensive development of biochemical and genetic methods to distinguish taxa (6), evolutionary relationships among genotypes of extant *Trichinella* remain enigmatic. Delineating the phylogeny and biogeography of *Trichinella* will substantially increase the comparative value of its genome sequence and help elucidate the determinants of zoonotic risk among species. On a broader scale, this information will provide us an opportunity to investigate ecological and evolutionary forces that promote diversification in complex host-parasite systems.

Three central questions fueled our curiosity about the evolution of *Trichinella*. (i) Can prolonged coevolution with distinct tetrapod groups explain why encapsulated species of *Trichinella* are restricted to eutherian mammals and nonencapsulated species also infect saurian, crocodilian, avian, eutherian, and metatherian hosts (5, 7), or have more recent ecological and biogeographic events influenced current host associations? (ii) Why are most species and genotypes with broad host ranges

Conflict of interest statement: No conflicts declared.

Abbreviations: SSU, small-subunit; LSU, large-subunit; ITS-2, second internal transcribed spacer; ML, maximum likelihood; MY, million years; COI, cytochrome oxidase I.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY851256–AY851277 and DQ007890–DQ007900).

[§]To whom correspondence should be addressed at: U.S. Department of Agriculture, Agricultural Research Service, Bovine Functional Genomics Laboratory, B1180 Beltsville Agricultural Research Center East, Beltsville, MD 20705. E-mail: zarlenga@anri.barc.usda.gov.

© 2006 by The National Academy of Sciences of the USA

Table 1. Characteristics of sequence alignments and their evolutionary models

Data set*	No. of bases [†]			Model type [‡]	C:T/Tv	G:A/Tv	Base frequencies			I	I' shape	Clock [§]	
	Total	Var.	Inf.				A	C	G			ΔlnL	df > P
SSU	1,929	717	417	TrNefI [†]	2.80	2.06	0.250	0.250	0.250	0	0.408	20.3	14 > 0.12
<i>11mit</i>	1,193	276	194	HKY I [†]	15.60	15.60	0.299	0.170	0.174	0.49	0.492	9.84	9 > 0.35
<i>3gene</i>	1,273	204	78	HKY I [†]	3.21	3.21	0.277	0.165	0.200	0	0.156	1.86	6 > 0.93

*Data sets are as follows: SSU, 10 of 11 *Trichinella* genotypes including outgroup taxa *Trichuris muris*, *Trichuris suis*, *Adnoncholaimus* sp., *Pirapulus caudatus*, *Mormis nigrescens*, and *Mylonchulus arenicolu*; 11mit, mitochondrial LSU rDNA and COI from all 11 *Trichinella* taxa; 3gene, mitochondrial LSU rDNA, COI, and ITS-2 from all eight encapsulated species and genotypes of *Trichinella*.

[†]Total, variable (Var.), and informative (Inf.) number of bases under the parsimony criterion.

*TnF, a model consisting of a single transversion rate but two distinct transition rates. HKY models correspond to that of Hasegawa *et al.* (8) in assuming unequal base frequencies and rates of transition and transversion substitutions. Γ , variation in the substitution across nucleotide sites modeled as a gamma distribution with an estimated shape parameter; I, the estimated proportion of invariant sites; Tv, transversion.

⁵Difference in log likelihood ($\Delta\ln L$) trees evaluated with and without the enforcement of a molecular clock. $df = (\text{number of taxa}) - 2$. P represents statistical significance in the reduction of likelihood owing to enforcement of a molecular clock.

restricted to well defined geographical localities? (iii) Did humans become infected with *Trichinella* in the Neolithic before domesticating pigs, or did adaptation after pig domestication cause certain species of *Trichinella* to threaten human health? To explore these and related questions, we generated the first robust hypothesis for the phylogeny of *Trichinella* based on multilocus DNA sequence information from all ecologically and genetically recognized species and genotypes. We then interpreted these results with reference to the evolutionary history and biogeography of their hosts.

Results

The general features of the three comparative data sets are summarized in Table 1. Significant variation in the rate of evolution across sites was identified in each model. In no case did enforcing a molecular clock significantly worsen the likelihood of the data.

An initial analysis of the SSU rDNA variation served to evaluate *Trichinella* phylogeny within the broader context of nematode evolution. In this regard, the SSU rDNA sequences reflect a minimally differentiated, monophyletic assemblage (Fig. 1) where the maximum likelihood (ML) pairwise distance never exceeded 0.0052 among encapsulated species or 0.0135

among all species. These values are substantially less than the ML distance (0.2) separating *Trichuris muris* from *Trichuris suis*. If the locus evolved in each lineage at $\approx 0.4 \times 10^{-9}$ per site per year (4% over 100 MY), which typifies many metazoan phyla (9), then extant species of *Trichinella* shared a common ancestor ≈ 16 million years (MY) ago, and encapsulated taxa first began diversifying within the last 9 MY (Table 2). Only by postulating an implausible 18-fold rate decrease in divergence rate could the extant diversity among *Trichinella* species date to the origins of distinct tetrapod host groups that occurred some 300 MY B.P. Although species of *Trichinella* are poorly differentiated from one another at the SSU rDNA, this locus markedly distinguishes them from species of *Trichuris*. The extant pairwise difference between *Trichinella* and *Trichuris* exceeds 16%, and ML estimates of their genetic distance exceed 22%. Assuming a per lineage rate of 4% over 100 MY for SSU rDNA evolution, the most recent common ancestor of *Trichinella* and *Trichuris* may have existed 275 MY B.P.

Diverse tree-building approaches yielded consistent and nearly complete phylogenetic resolution among all 11 *Trichinella* congeners when compared at the two mtDNA genes (Fig. 1), or among all eight encapsulated species when the second internal transcribed spacer (ITS-2) sequences were also included (data

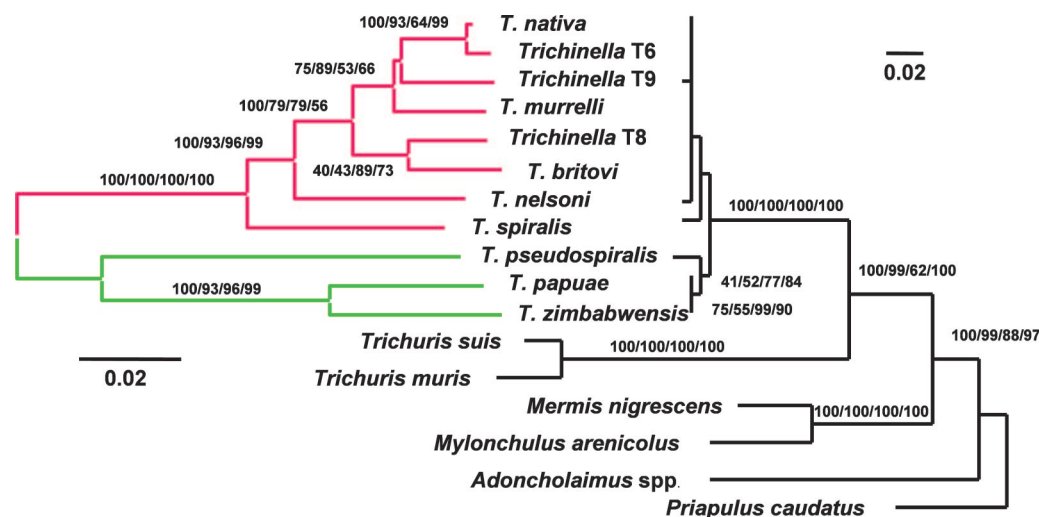


Fig. 1. Midpoint rooted minimum evolution trees reconstructed from all known encapsulated (red) and nonencapsulated (green) species and genotypes of *Trichinella* based on the variation in mitochondrial LSU and COI DNA (on the left) and SSU rDNA (on the right). Topological support is indicated by Bayesian posterior probabilities and by ML, minimum evolution (using ML distances), and parsimony bootstrap replicate analyses (B/ML/ME/P). Bootstrap support was reconstructed from 100-bp replicates.

high-latitude Holarctic (-5°C January isotherm) during the Wisconsin or terminal Pleistocene. Indeed, secondary contact zones in western North America have developed for some hosts tracking the retreat of glaciers (29). Populations of *T. nativa* may have expanded into high arctic habitats of Canada and Greenland during the Holocene, as has been postulated for other Holarctic parasitic nematodes and their mammalian hosts (26, 30, 31). Alternatively, assuming that peripheral isolation played a role in diversifying these crown species, *T. nativa* may instead represent an ancestral and persistent high-latitude population in the Holarctic.

To date, no endemic species of *Trichinella* has been identified in the Neotropics even though felids, procyonids, canids, and mustelids colonized South America after formation of the Panamanian Isthmus during the Pliocene/Pleistocene ≈ 3.0 MY B.P. (32). In South America only *T. spiralis* has been found, presumably the result of anthropogenic translocation. Among sylvatic *Trichinella* adapted for survival in harsh, cold, and xeric environments, historical constraints and life history canalization may have precluded *Trichinella* from expanding into the Neotropics across the Panamanian Isthmus.

Human Trichinellosis

Did the divergence of *T. spiralis* coincide with the domestication of pigs? If so, its independent history would have commenced no earlier than the Neolithic, and a shallow genetic divergence would be evident between the parasites of wild and domesticated suids. Instead, *T. spiralis* appears to have commenced its independent evolutionary trajectory several MY before *Sus scrofa* was first domesticated (33) because none of the enzootic species share an especially close relationship with *T. spiralis*. Thus, the ecological setting in which hominids first acquired trichinellosis may parallel that of *Taenia* tapeworms (34) and lice (35). Hunting prey also pursued by felids, canids, and hyaenids may have exposed early humans to a suite of parasites already well established on the savannahs of Africa during the late Tertiary and Quaternary. Accordingly, hominids would have first acquired *Trichinella* when their diets shifted from herbivory to scavenging and facultative carnivory during the Pliocene or early Pleistocene, (17, 36–38).

Recent host adaptations may have enhanced the risk to human health posed by certain species of *Trichinella*; however, a broad host range characterizes basal lineages and therefore represents the ancestral condition. The capacity of nonencapsulated forms to infect a broad array of terrestrial vertebrates suggests they pose a significant zoonotic risk as well (11, 38, 39). By contrast, the capacity to infect swine is diminished among more recently radiated species (*T. britovi* and *T. nelsoni*) and is nonexistent among the crown species (*T. nativa*, *T. murrelli*, *Trichinella* T6, and *Trichinella* T9) (5).

Conclusions

Although the Trichinellidae and the Trichuridae apparently diverged as early as the Paleozoic, contemporary species of *Trichinella* diversified during the age of modern Eutheria. The following observations support this line of reasoning: (i) the genetic diversity among extant species of *Trichinella* accumulated within the last 15–20 MY; (ii) all extant species infect Eutherian carnivores or rodents; and (iii) infections in saurians, crocodilians, metatherians, and birds appear secondary rather than relictual. If, as we contend, extant diversity originated in Eurasia and disseminated to both Africa and the Nearctic during the late Tertiary, *Trichinella* should resemble other parasites for which host switching rather than strict cospeciation was a major diversifying force (40). These patterns underscore the role of episodic expansion, isolation, and host switching in promoting *Trichinella* diversification and highlight the exceptional dissemination potential conferred to *T. spiralis* by its association with domesticated swine and synanthropic rodents.

Table 3. *Trichinella* strains used in the analysis

Parasite	Genotype	ISS code
<i>T. spiralis</i>	T1	ISS003
<i>T. nativa</i>	T2	ISS010
<i>T. britovi</i>	T3	ISS005
<i>T. murrelli</i>	T5	ISS035
<i>Trichinella</i> T6*	T6	ISS040
<i>T. nelsoni</i>	T7	ISS029
<i>Trichinella</i> T8*	T8	ISS124
<i>Trichinella</i> T9*	T9	ISS409
<i>T. pseudospiralis</i> [†]	T4	ISS013
<i>T. papuae</i> [‡]	T10	ISS572
<i>T. zimbabwensis</i> [‡]	T11	ISS1029

ISS, Istituto Superiore di Sanità. Additional information can be obtained from www.iss.it/site/trichinella/.

*Genetically distinct but taxonomically undefined taxa.

[†]Nonencapsulated species infectious for birds in addition to mammals.

[‡]Nonencapsulated species infectious for reptiles in addition to mammals.

Materials and Methods

All species and genotypes of *Trichinella* (Table 3) were maintained in Swiss–Webster mice. Eviscerated mouse carcasses were digested with pepsin (1%) and HCl (1%), and the released muscle larvae were collected and washed through multiple exchanges in water. Genomic DNA was isolated by proteinase K and SDS digestion and treated with RNase. At least three full-length clones of PCR-amplified SSU rDNA were sequenced from each species and genotype by using *Trichinella*-specific primers (GenBank accession no. U60231). PCR-amplified mitochondrial large-subunit (LSU) rDNA and cytochrome oxidase I (COI) genes were directly sequenced. Three multiple sequence alignments were constructed by using CLUSTAL-X and visually confirmed. The first data set (18S) aligned the SSU rDNA of *Trichinella* species to homologues from several outgroup taxa (Table 1). To obtain better resolution of intragenetic relationships, we analyzed a second data set (11-mit) incorporating two mtDNA loci (COI and LSU rDNA) from all 11 species and genotypes of *Trichinella*. Inasmuch as the ITS-2 sequences of the eight encapsulated *Trichinella* taxa could not be aligned reliably with those of the nonencapsulated species, only the interrelationships among encapsulated forms were further explored by using a third data set (3-gene) by combining the ITS-2, COI, and mitochondrial LSU rDNA gene sequences. Likelihood ratio tests were used to identify which of 56 nucleotide substitution models best fit each of these three data sets by using MODEL TEST 3.0.4 (41) (Table 1) and to evaluate the compatibility of each data set with a molecular clock (42).

Phylogenies were reconstructed from bootstrap replicates of these alignments under the criteria of minimum evolution and ML by using the optimized model as implemented by PAUP* (43). Additionally, a Bayesian approach was used to simultaneously optimize tree topology and nucleotide substitution parameters, including six independent nucleotide substitution rates, the proportion of invariant sites, the shape parameter of the gamma distribution, and the frequency of each nucleotide, by using MR. BAYES 3.0 (44). Of a total 10 million generations, 1 of every 1,000 was sampled after discarding a “burn in” period of 100,000 generations. ML searches performed under topological constraints were also used to evaluate the relative statistical support of competing topologies.

Genetic distances were used to estimate the ages of divergence among species and genotypes of *Trichinella*. Several approaches were considered for each data set to derive a temporal range. For SSU rDNA, a per-lineage calibration rate of 0.4×10^{-9} per site per year was assumed based on data generally supported for metazoan phyla (9). For mtDNA, data

